



UNIVERSITY  
OF TAMPERE

This document has been downloaded from  
Tampub – The Institutional Repository of University of Tampere

Tekijät: Ahlgren Per, Järvelin Kalervo  
Artikkelin nimi: Measuring Impact of 12 Information Scientists Using the DCI index  
Julkaisuvuosi: 2010  
Lehden nimi: Journal of the American Society for Information Science and Technology (JASIST)  
Lehden vol: 61  
Lehden numero: 7  
Sivunumerot: 1424-1439  
ISSN: 1532-2882  
Tieteenala: Luonnontieteet / Tietojenkäsittelyoppi ja informaatiotieteet  
Kieli: en  
Yksikkö: Informaatiotieteiden yksikkö  
URN: <http://urn.fi/urn:nbn:uta-3-476>  
DOI: <http://dx.doi.org/10.1002/asi.21333>

All material supplied via TamPub is protected by copyright and other intellectual property rights, and duplication or sale of all part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorized user.

# Measuring Impact of 12 Information Scientists Using the DCI index

Per Ahlgren<sup>a</sup> and Kalervo Järvelin<sup>b</sup>

<sup>a</sup> Department of e-Resources, University Library, Stockholm University, Sweden

<sup>b</sup> Department of Information Studies and Interactive Media, University of Tampere, Finland

email: per.ahlgren@sub.su.se; kalervo.jarvelin@uta.fi

Corresponding author: Per Ahlgren, Department of e-Resources, University Library, Stockholm University, 106 91 Stockholm, Sweden

## Abstract

The Discounted Cumulated Impact (DCI) index has recently been proposed for research evaluation. In the present work, an earlier data set by Cronin & Meho (2007) is reanalyzed, with the aim of exemplifying the salient features of the DCI index. We apply the index on, and compare our results to the outcomes of, the Cronin-Meho (2007) study. Both authors and their top publications are used as units of analysis, which suggests that, by adjusting the parameters of evaluation according to the needs of research evaluation, the DCI index delivers data on an author's (or publication's) 'lifetime' impact or current impact at the time of evaluation, on an author's (or publication's) capability of inviting citations from highly-cited later publications as an indication of impact, and on the relative impact across a set of authors (or publications) over their 'lifetime' or currently.

Keywords: Citation impact, DCI-index, performance evaluation, research evaluation

## Introduction

*Bibliometrics*, which may be defined as the quantitative study of publication collections, can be used to illuminate the scientific impact of different units. The impact of countries, institutions, departments and individual researchers can be measured on the basis of the number of citations the corresponding publications have received (e.g., Glänzel, Debackere, & Meyer, 2008; Meho & Rogers, 2008; Takeda, Mae, Kajikawa, & Matsushima, 2009; Visser & Nederhof, 2007; Zhou, Thijs, & Glänzel, 2009). Such an approach may be regarded as a complement to the well-known peer-review approach, where the evaluation of a given unit is exclusively based on peer judgment. There are at least two reasons for complementing peer-review with citation impact. Firstly, the citation impact approach provides a global perspective: the researchers that cite a given unit may come from any country of the world. One might say, with an overstatement, though, that the (implicit) evaluation is performed worldwide. This contrasts with the local perspective provided by peer-review. Secondly, peer-review does not always work as unproblematic as is sometimes assumed. For instance, nepotism in connection with peer-review has been reported in the literature (Moed, 2005; Sandström & Hällsten, 2008; Wennerås & Wold, 1997).

As public and private research funding entities, like universities, government offices and boardrooms, need to assess the quality of applicants and the effects of their investments,

bibliometric research evaluation is an area of increasing importance (Visser & Nederhof, 2007). Several measures are utilized in this kind of evaluation. A simple and popular measure is the *h*-index (Hirsch, 2005), a measure that combines citation impact and productivity. The index has been used to rank researchers (e.g., Cronin & Meho, 2006; Oppenheim, 2007), as well as calculated for journal citations (Braun, Glänzel, & Schubert, 2005; Egghe & Rousseau, 2006; Schubert & Glänzel, 2007). However, the *h*-index has drawbacks, and variations of the index have been proposed to overcome some of them (Burrell, 2007; Egghe, 2006; Jin, Liang, Rousseau, & Egghe, 2007).

A more sophisticated measure than the *h*-index, with regard to citation impact, is the *field normalized citation rate* (Braun & Glänzel, 1990; Moed, 2005). For a given unit of analysis, the mean number of citations per publication is calculated. The resulting number is then divided with a summarizing expected value regarding publications (a) of the same type (article, review, and so on), (b) from the same year, and (c) published in journals belonging to the same field, as the publications from the unit. This (relative) measure, which utilizes the ISI/Thomson Reuters classification of journals into a large number of subject classes (fields), is such that units with different subject profiles can be compared in a meaningful way.

The Discounted Cumulated Impact (DCI) index was recently proposed as a research evaluation tool (Järvelin & Persson, 2008). It is based on the idea of devaluing old citations in a smooth and parameterizable way. It rewards a unit of analysis for receiving new citations even if the cited publication is old. Further, it allows weighting of the citations by the citation weight of the citing publication. The DCI index can be used to calculate research performance on the basis of the *h*-core of a researcher, or any other publication data set for a researcher, like all publications included in the Web of Science. While the study by Järvelin & Persson (2008) proposed the DCI index, and exemplified its functioning through a small empirical case, serious empirical applications of the index are still to be seen.

The present work is a serious empirical application of the DCI index. It builds on an earlier work by Cronin & Meho (2007) who analyzed the citation impact of 12 authors. Each author is such that the author has won either the American Society for Information Science and Technology Award of Merit or the Research in Information Science award, or both (for background information on these awards, see: <http://www.asis.org/awards/winners.html>). All 12 authors – Marcia Bates, Nicholas Belkin, Blaise Cronin, Raya Fidel, Paul Kantor, Carol Kuhlthau, Gary Marchionini, Tefko Saracevic, Dagobert Soergel, Don Swanson, Carol Tenopir, and Howard White – are recognized information scientists.

The goal of the present work is to show salient features of the DCI index in citation impact analysis, through applying it on, and comparing its results to the outcomes of, the Cronin-Meho (2007) study. We shall use both authors and their top publications as units of analysis. In particular, we aim at showing that, by adjusting the parameters of evaluation according to current needs of research evaluation, the DCI index delivers data on:

- a unit's 'lifetime' impact or current impact at the time of evaluation
- a unit's capability of attracting citations from highly-cited later publications as an indication of impact, and
- the relative impact across a set of units, also regarding their 'lifetime' or current impact.

This can reveal rapidly rising new stars in a field as well as the units, which are becoming part of the history of a field – no more actively cited. Traditional analyses based on total

accumulated citations, or the  $h$ -index, do not directly support this, and require special treatment of the data sets used. By using the term ‘unit’, we refer to the fact that DCI-based evaluation can be applied at any aggregation level from individual publications to scholars, institutions, research fields, and countries.

We point out that the present work is a case study, and not a big sample study involving comprehensive comparisons with other indicators. Its findings might inspire other researchers to further investigate, empirically or theoretically, the properties of the DCI index.

The paper is organized as follows. We first present the data and methods employed in this study. The methods are divided into overall author impact analysis of the 12 authors and their top publication impact analysis. The section on findings reports these analyses graphically. Finally we compare our findings to the earlier ones and seek to point out the salient features of the DCI type-of-analysis.

## Data and Methods

This study has two parts. One part has the 12 authors as units of analysis, whereas the other has selected publications of the authors as units of analysis. For both parts, we throw light upon the impact of the units of analysis in terms of received citations. Before we describe the two parts, we deal with the research performance index used in the study and with weighting of citing publications.

### Discounted Cumulated Impact

As indicated above, we apply a recent research evaluation tool: the *DCI index* (Järvelin & Persson, 2008). Let  $\mathbf{F} = (x_1, x_2, \dots, x_t)$  be a vector such that each component  $x_i$  is non-negative. A given  $x_i$  is the impact value the unit under evaluation has received in the  $i$ th period of  $t$  considered periods. We call vectors of this kind *impact vectors*. Let  $j$  ( $2 \leq j \leq t$ ) be the current period of evaluation. We define the *discounted cumulated impact vector* of length  $j$ ,  $\mathbf{DCI}_j$ , recursively as:

$$\mathbf{DCI}_j[i] = \begin{cases} x_1 / \max\{1, \log_b(j-1)\} & \text{if } i = 1 \\ \mathbf{DCI}_j[i-1] + x_i / \max\{1, \log_b(j-i)\} & \text{if } 1 < i < j \\ \mathbf{DCI}_j[i-1] + x_i & \text{if } i = j \end{cases} \quad (1)$$

where  $\mathbf{DCI}_j[i]$  denotes the  $i$ th component in  $\mathbf{DCI}_j$ . When  $j = 1$ , we let  $\mathbf{DCI}_1[1]$  be equal to  $x_1$  (the first component in  $\mathbf{F}$ ), which yields  $\mathbf{DCI}_1 = (x_1)$ . The approach here is such that old citations (with small  $i$  and thus large discount component  $\log_b(j-i)$ ) are more heavily discounted than newer ones (with large  $i$ ). We note, further, that no discount occurs when  $i = j$ , that is,  $x_j$  is added without any discount to the cumulated value. Actually, neither  $x_{j-1}$  is reduced because  $\max\{1, \log_b(j-(j-1))\}$  yields 1. Sharper or smoother discounts can be obtained by varying the value of  $b$ . For instance, a low value on  $b$  yields a sharp discount, and would model a short collective memory.  $b$ , then, can be regarded as an impact value decay parameter.

We now turn our attention to the *dynamic discounted cumulated impact vector* of length  $k$ ,  $\text{dDCI}_k$ , where  $1 \leq k \leq t$ . This vector, which cumulates and discounts past performance at each period of evaluation up to period  $k$ , is defined as:

$$\text{dDCI}_k[j] = \text{DCI}_j[j] \quad (2)$$

for each  $j$  such that  $1 \leq j \leq k$ . Thus, the  $\text{dDCI}_k$  vector is defined in terms of  $k$   $\text{DCI}$  vectors. For example, if  $t = 20$  and  $k = 10$ , then  $\text{dDCI}_{10}[5] = \text{DCI}_5[5]$ . That is, the fifth component in  $\text{dDCI}_{10}$  is equal to the fifth and last component in  $\text{DCI}_5$ , the cumulated value at period 5 when period 5 is the current period of evaluation.

## Normalized Dynamic Discounted Cumulated Impact

Consider a set  $V = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  of dynamic discounted cumulated impact vectors, where each  $\mathbf{v} \in V$  is of length  $k$ . Let  $\mathbf{m}$  be the vector  $(-1, -1, \dots, -1)$ , where the number of components is  $k$ , and let  $\mathbf{w} \in V$ . We define the *normalized dynamic discounted cumulated impact vector of length  $k$  with respect to  $\mathbf{w}$  and  $V$* ,  $\text{dDCIn}_{k(\mathbf{w}, V)}$ , as

$$\text{dDCIn}_{k(\mathbf{w}, V)} = \left( \mathbf{w} / \frac{1}{n} \sum_{\mathbf{v} \in V} \mathbf{v} \right) + \mathbf{m} \quad (3)$$

The division operation (“/”) is performed component by component: each component in  $\mathbf{w}$  is divided by the corresponding component in the average vector for  $V$ , where the average vector is given by  $\frac{1}{n} \sum_{\mathbf{v} \in V} \mathbf{v}$ . This results in a new vector such that each component expresses the standing of its counterpart in  $\mathbf{w}$  relative to the average (across the  $n$  vectors and for the vector position of the component). To this relative-to-the-average vector  $\mathbf{m}$  is added, which yields that 1 is subtracted from each component in the former vector, thereby rendering average impact as 0 and total neglect as  $-1$ .

## Weighting of Citing Publications

For weighting of citing publications, we followed Järvelin & Persson (2008) and used the citation count of the citing publication for weighting. Thereby, we take 2<sup>nd</sup> generation citations into account<sup>1</sup>. However, we do not use the raw counts. Instead, these are normalized with the aid of logarithms as follows. Let *c-count* be the citation count of the citing publication and *cit* ( $1 < \text{cit} \leq 1000$ ) the base of the logarithm used to normalize citation counts. We define the *citing weight* of the citing publication, *c-weight*, as

$$c\text{-weight} = \begin{cases} 1 & \text{if } c\text{-count} = 0 \\ \max\{1, \log_{\text{cit}} c\text{-count}\} & \text{otherwise} \end{cases} \quad (4)$$

The minimal value of *c-weight* is 1. When *c-count*  $>$  *cit*, *c-weight* exceeds 1. When *cit* is small, say 2, one hundred citations gives a weight close to 7. If we set *cit* to 1000, practically

---

<sup>1</sup> As an example, 2<sup>nd</sup> generations citations are utilized in InCites, a new citation-based evaluation tool from Thomson Reuters. However, this tool does not use 2<sup>nd</sup> generation citations for weighting of citing publications.

all citing weights are equal to one ( $\log_{1000} 1000 = 1$ ). In this work, we let  $cit = 2$ . We did not apply weighting of citing publications transitively, for the following reasons. (1) We already take a log of the citation counts. A log of log of the transitive ones would have marginal impact. (2) Due to time lags in citing, a first generation citation (being older) is anyway discounted, so the effect of transitivity is further reduced. (3) The significance of transitive citing is not well-understood, since it disperses across the set of references of the first generation citing publication.

## Authors as Units of Analysis

For the first part of the study, we used an extension, say  $L_{citP}$ , of the list of citing publications used by Cronin & Meho (2007).  $L_{citP}$  was obtained by extending the Cronin-Meho list with citing publications gathered in February 2009, and published too late to be included in the study by Cronin & Meho<sup>2</sup>. Let  $A$  be one of the 12 authors. Each publication in the sublist of  $L_{citP}$  associated with  $A$  is such that it cites at least one publication, which has  $A$  as one its authors. A publication in the sublist associated with  $A$  is taken to assign exactly one citation to  $A$ , irrespective of how many publications by  $A$  the publication cites. Therefore, the *citation frequency* for  $A$  with respect year  $k$  is taken to be the number of publications in the sublist of  $L_{citP}$  associated with  $A$  that are published year  $k$ .<sup>3</sup> The citation counts of the publications in the sublist associated with  $A$  were obtained from Web of Science 2009-06-05.

The total citation frequencies for the 12 authors are given in Table 1. For each author, it holds that the author's citations are counted from the date of the author's first citation in Web of Science.

**Table 1. Total citation frequencies for the 12 authors.**

Cited author	February 2009
Bates, Marcia J.	1039
Belkin, Nicholas J.	1344
Cronin, Blaise	1240
Fidel, Raya	648
Kantor, Paul B.	802
Kuhlthau, Carol C.	756
Marchionini, Gary	1037
Saracevic, Tefko	1603
Soergel, Dagobert	478
Swanson, Don R.	975
Tenopir, Carol	1023
White, Howard D.	781

For each author, we constructed two impact vectors,  $F_{au}$  and  $F_{au(w)}$ , both of length 54 and such that a given position  $i$  corresponds to a publishing year (of citing publications). Position 1 corresponds to 1956, the first year  $k$  such that at least one publication, published year  $k$ ,

<sup>2</sup> The new citing publications were gathered by Lokman Meho, who also extended the Cronin-Meho list.

<sup>3</sup> Clearly, this approach is not optimal. However, the approach was applied by Cronin & Meho (2007), and we wanted to follow these authors closely for comparative reasons.

cites at least one the 12 authors. The last position corresponds to 2009. For  $\mathbf{F}_{au}$  holds that a given component,  $x_i$ , is the citation frequency for the author with respect to the year that corresponds to position  $i$ . Regarding  $\mathbf{F}_{au(w)}$ ,  $x_i$  is the sum of c-weights of the publications, which give rise to the citation frequency for the author with respect to the year that corresponds to position  $i$ .

From the  $(2 \times 12)$  impact vectors, we generated, for both the non-weighting and the weighting case, dynamic discounted cumulated impact vectors (of length 54). Moreover, we used three values on the decay parameter  $b$ , for both cases, to see the effect of the parameter on the DCI index. The three values used, 1.1, 2 and 10, corresponds to rapid, moderate and slow decay of past impact, respectively. With a large value on  $b$ , old citations keep all, or almost all, of their weight. Therefore, when  $b$  is large, an author's past citation history plays an equally, or almost equally, important role as the more recent history, with respect to the DCI impact at a given evaluation time. With large values on  $b$ , the overall impact of an author is indicated, in the sense that also the old citations are taken into account, with equal, or almost equal weight as the recent ones. With a value on  $b$  that is close to 1 (but greater than 1), it is the recent history that has almost all the importance in evaluation.

Finally, from the dynamic discounted cumulated impact vectors, we generated normalized dynamic discounted cumulated impact vectors (of length 54) across the set of our 12 authors. These latter vectors were used in the author analysis part of the study. Thus, we evaluated the impact of the authors under six different scenarios (Table 2).

**Table 2. Authors as units of analysis. The six evaluation scenarios.**

Scenario	Type-of-Analysis
au-1	Normalized non-weighted author dDCI with decay $b = 1.1$
au-2	Normalized non-weighted author dDCI with decay $b = 2$
au-3	Normalized non-weighted author dDCI with decay $b = 10$
au-4	Normalized citation-weighted author dDCI with decay $b = 1.1$
au-5	Normalized citation-weighted author dDCI with decay $b = 2$
au-6	Normalized citation-weighted author dDCI with decay $b = 10$

## Publications as Units of Analysis

For the second part of the study, we used Web of Science in order to obtain the first order citations to the top publications, and the second order citations to these citing publications as well. We selected the top publications from among the high-impact works studied by Cronin & Meho (2007). For each author, except for Soergel, we selected the most highly cited publication of the author among the ones that are listed in the appendix in Cronin & Meho (Cronin & Meho, 2007). Soergel was left out, since only books are listed for this author, and citations to books are not recorded in Web of Science. The 11 selected top publications are listed in Appendix. For each selected publication, citing publications, and their citation counts, were obtained on 2009-08-31.

We constructed, for each publication, two impact vectors,  $\mathbf{F}_{pub}$  and  $\mathbf{F}_{pub(w)}$ , both of length 46 and such that a given position  $i$  corresponds to a publishing year (of citing publications). Position 1 corresponds to 1963, the year in which the oldest of the 11 publications was published, while the last position corresponds to 2009.  $\mathbf{F}_{pub}$  is such that a given component,

$x_i$ , is the number of publications, published in the year that corresponds to  $i$ , that cite the publication. For  $\mathbf{F}_{pub(w)}$ ,  $x_i$  is the sum of c-weights of the publications, which are published in the year that corresponds to  $i$  and which cite the publication.

**Table 3. Publications as units of analysis. The 12 evaluation scenarios.**

Scenario	Type-of-Analysis
p-1	Non-weighted publication DCI with decay $b = 1.1$
p-2	Non-weighted publication DCI with decay $b = 2$
p-3	Non-weighted publication DCI with decay $b = 10$
p-4	Weighted publication DCI with decay $b = 1.1$
p-5	Weighted publication DCI with decay $b = 2$
p-6	Weighted publication DCI with decay $b = 10$
p-7	Non-weighted publication dDCI with decay $b = 1.1$
p-8	Non-weighted publication dDCI with decay $b = 2$
p-9	Non-weighted publication dDCI with decay $b = 10$
p-10	Weighted publication dDCI with decay $b = 1.1$
p-11	Weighted publication dDCI with decay $b = 2$
p-12	Weighted publication dDCI with decay $b = 10$

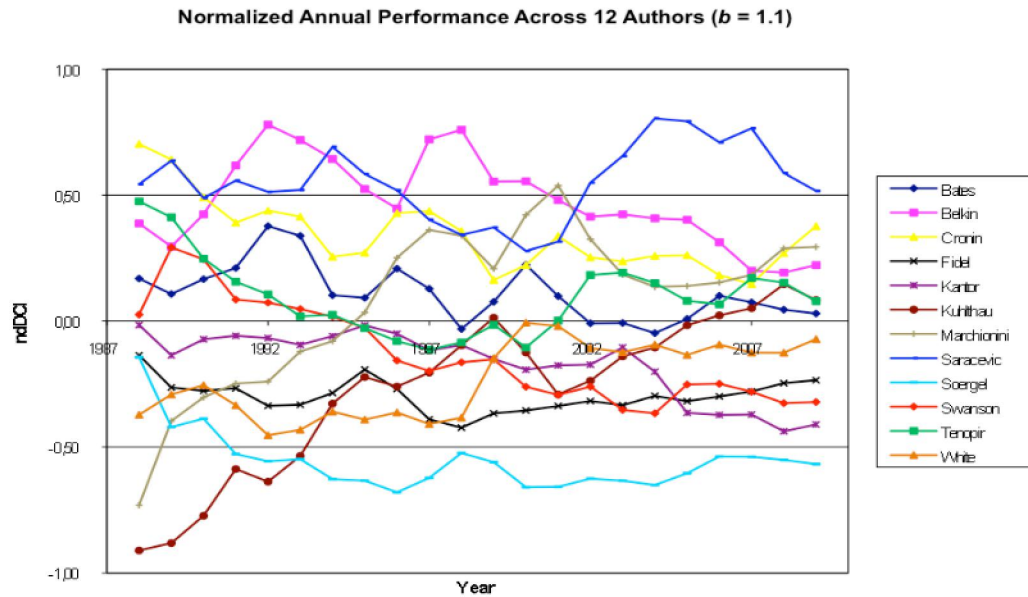
From the  $(2 \times 11)$  impact vectors, we generated, for both the non-weighting and the weighting case, (dynamic) discounted cumulated impact vectors (of length 46). In this part of the study, we again used 1.1, 2 and 10 as values on the decay parameter  $b$  to see the effect of the parameter on the DCI index. Normalized dynamic discounted cumulated impact vectors were not employed here (such an analysis is perhaps not relevant across authors, while it would be interesting within authors). Thus, we evaluated the impact of the publications under 12 different scenarios (Table 3).

## Findings

### Normalized Author Performance

We first look at the unweighted normalized dynamic citation performance of our 12 authors and vary the citation decay parameter from  $b = 1.1$  to 2 to 10. This represents how long a history we incorporate in citation performance with  $b = 1.1$  representing a very short history (i.e. ‘currency’) and  $b = 10$  allowing past performance to carry forwards in time. Recall that normalization here works between the 12 authors, zero-level representing their average performance. The graphs cover the period 1988 to 2009 for which we have data for all the authors.

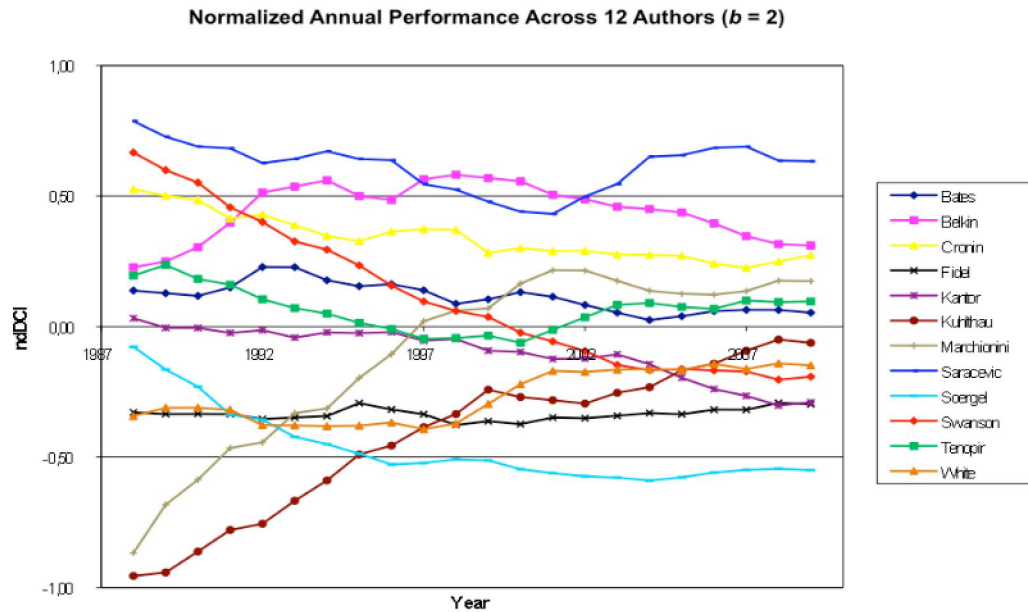




**Figure 1.** Normalized dynamic DCI of 12 authors with  $b = 1.1$ , no citation weights.

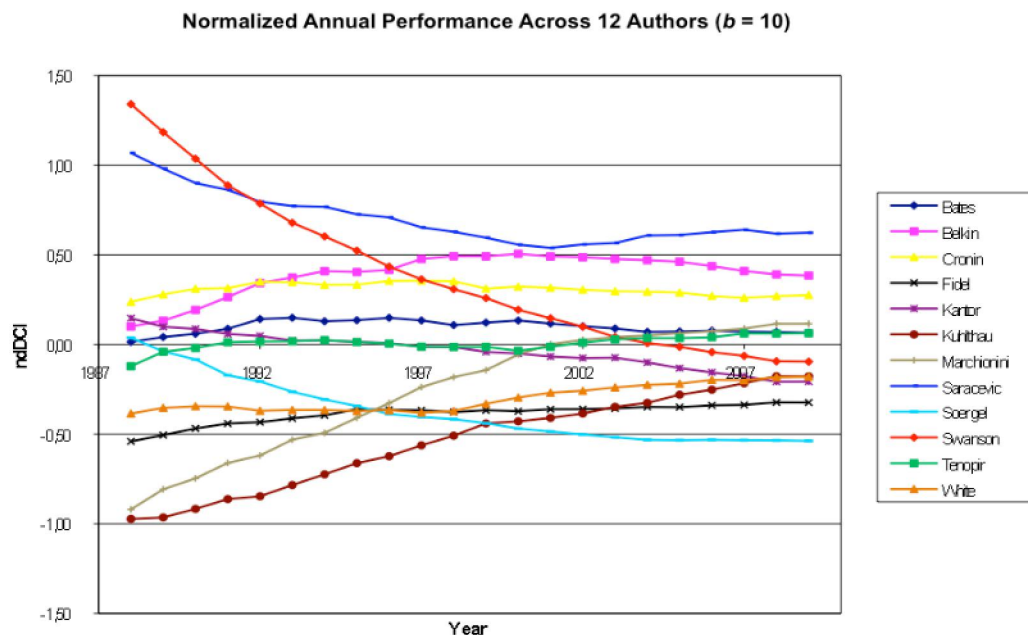
Figure 1 displays the unweighted normalized dynamic DCI of our 12 authors with decay  $b = 1.1$ . The individual curves are rather jagged, reflecting the annual variation in the number received citations and little historical smoothing. One may see that Marchionini and Kuhlthau, who were in 1988 in their early careers, in 10 year's time climb up to the average level and above. Marchionini is top scoring for one year, 1999. The three constantly top scoring scholars are Belkin, Cronin and Saracevic with Belkin at the top in the first half and Saracevic across the most recent years. Marchionini captures the fourth rank from Bates by mid-1990's.

Figure 2 visualizes corresponding data on the 12 authors with decay  $b = 2$ . The individual curves have become smoother because more of past performance is taken into account for each year of evaluation. The overall picture does not change greatly but there are interesting details. Saracevic strengthens his top position, Belkin the second, then come Cronin and Marchionini passing Bates in 2000. Swanson's late career is reflected in his curve dropping from rank 2 in 1988 to rank 9 in 2009.



**Figure 2.** Normalized dynamic DCI of 12 authors with  $b = 2$ , no citation weights.

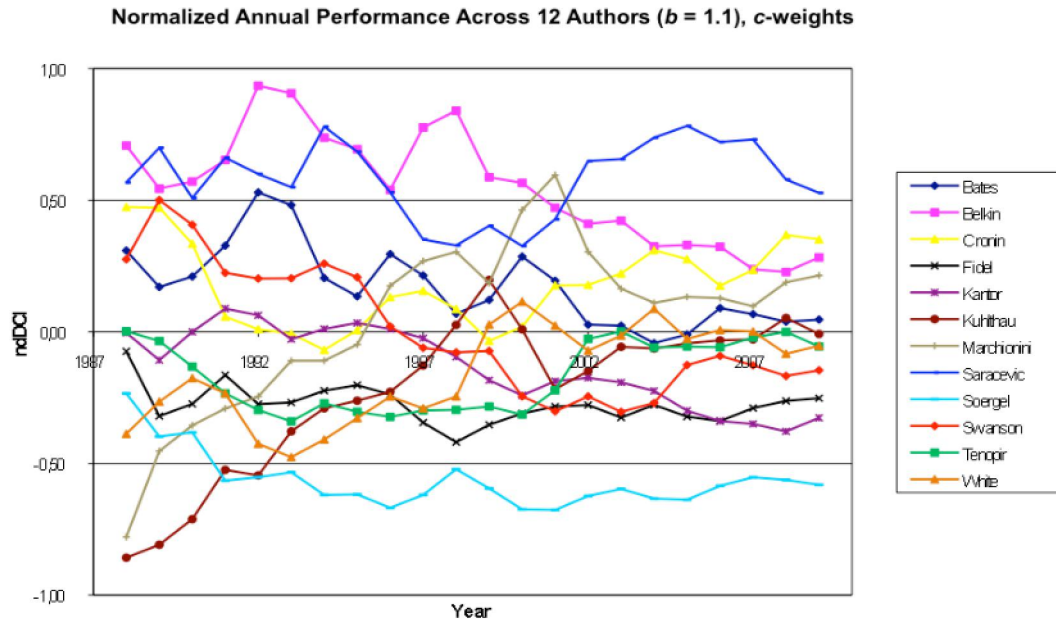
Figure 3 visualizes corresponding data on the 12 authors with decay  $b = 10$ . Now 10 years of citations are accumulated without discounts where after smooth discounting begins. This turns the individual curves even smoother. Swanson's strong performance prior to 1988 boosts him to top position for the first years. Saracevic, Belkin and Cronin secure their positions, and Bates gains momentum over Marchionini due to her longer career.



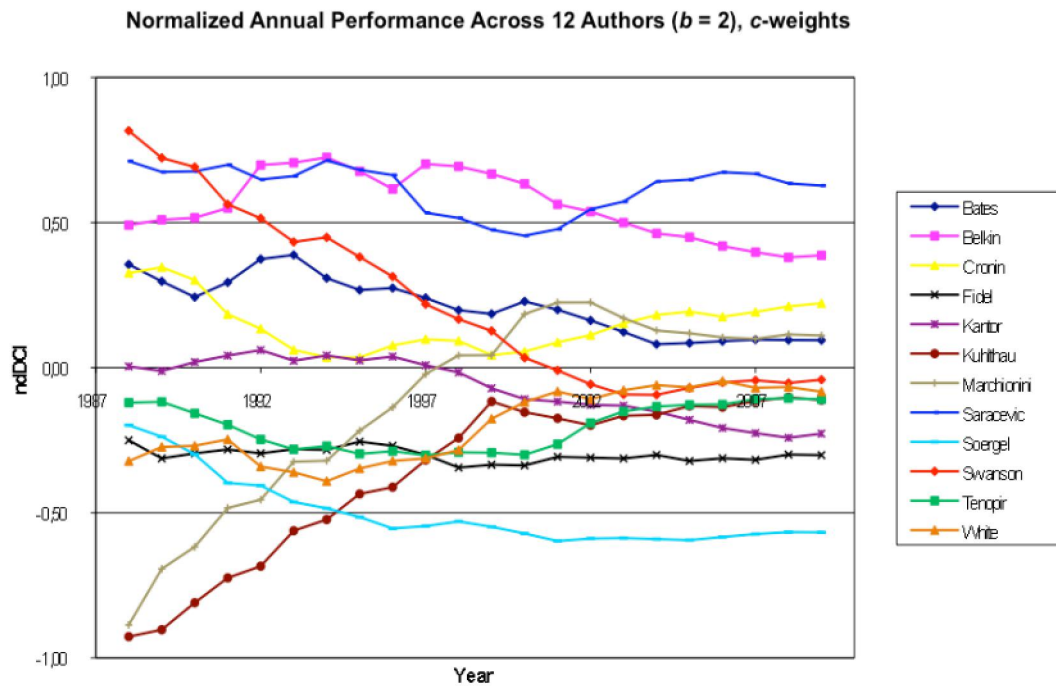
**Figure 3.** Normalized dynamic DCI of 12 authors with  $b = 10$ , no citation weights.

Across the Figures 1 to 3, we may observe that jagged curves turn smooth when the decay parameter  $b$  changes from 1.1 to 10. A small value of  $b$  yields curves that show normalized dynamic *currency* of authors whereas a large value, incorporating more historical impact, yields curves indicating each author's overall impact to a field. The longer the data series are, the more interesting become the graphs.

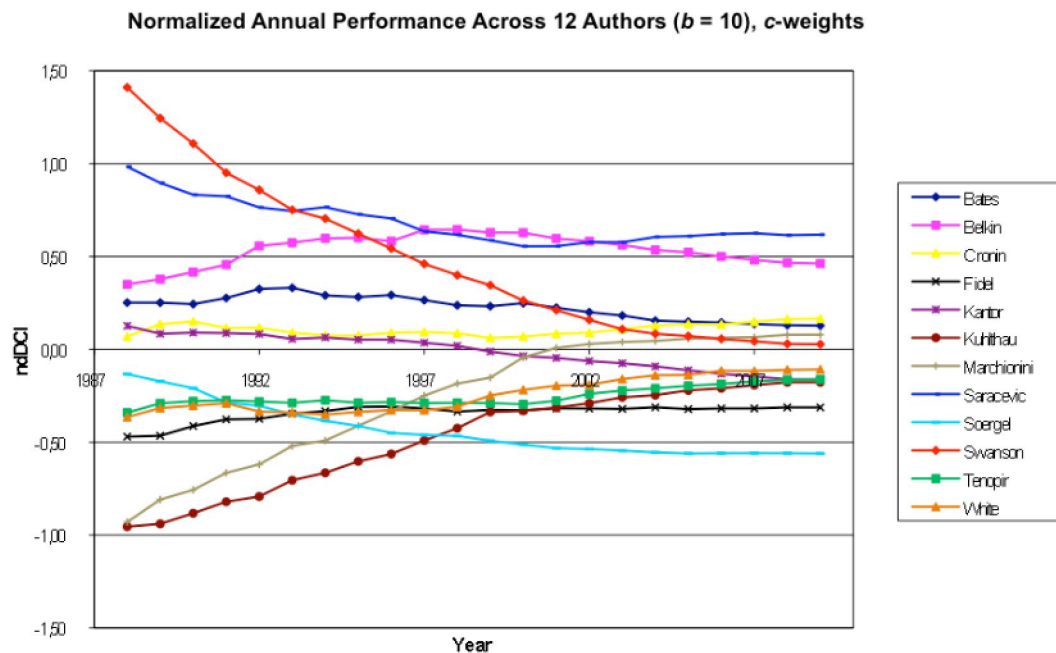
We now turn our attention to the *weighted* normalized dynamic DCI of our 12 authors using the same values on the decay parameter  $b$  as in the unweighted case (Figures 4 to 6).



**Figure 4.** Normalized dynamic DCI of 12 authors with  $b = 1.1$  and citation weights.



**Figure 5.** Normalized dynamic DCI of 12 authors with  $b = 2$  and citation weights.



**Figure 6.** Normalized dynamic DCI of 12 authors with  $b = 10$  and citation weights.

Figure 4 displays the weighted normalized dynamic DCI of our 12 authors with  $b = 1.1$ . The individual curves are again rather jagged, and the overall findings the same as in the unweighted case. Belkin and Saracevic are top scoring but the third position is not so clear. Marchionini and Kuhlthau quickly climb up to the average level and above, the former capturing the fourth rank. Comparing to Figure 1, weighting nevertheless suggests that in the first half of the period, Belkin was cited relatively more by high impact literature, while the opposite holds for Cronin. Weighting drops Tenopir's impact considerably and boosts Kantor's slightly.

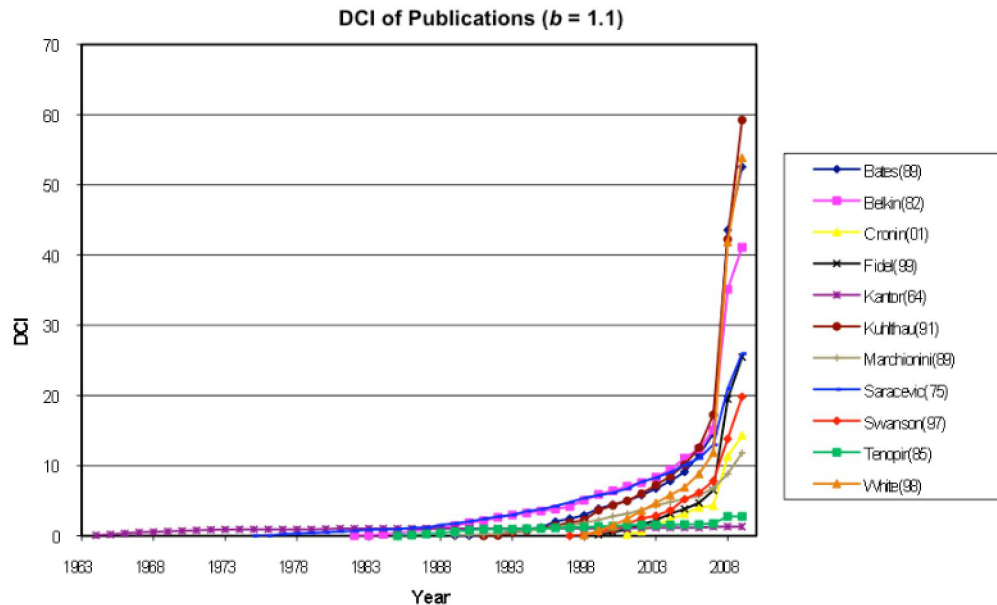
With a slower decay ( $b = 2$ ; Figure 5), Swanson's relative position is stronger, while Marchionini's and Kuhlthau's positions are slightly weaker. During the last half, Bates, Cronin and Marchionini have a nearly equal impact, ranked third. Comparing to Figure 2, also here weighting suggests that Belkin was initially cited more by high impact literature, while the opposite holds for Cronin. Weighting drops Tenopir's impact, and boosts Swanson's, considerably.

With even more historical impact included ( $b = 10$ ; Figure 6), Swanson's relative position grows even stronger, the other clear winners being Bates and Saracevic. Not surprisingly, it takes longer for Marchionini to reach the average impact among the authors while Kuhlthau does not quite make it at all in the analyzed time span. Comparing to the unweighted case (Figure 3), also here Belkin and Swanson gain impact, while the opposite holds for Cronin and Tenopir.

Figures 4 to 6 show that jagged curves turn smooth when the decay parameter  $b$  changes from 1.1 to 10. A small value on  $b$  reveals current weighted impact of authors for each year of analysis, while a large value reveals an author's overall weighted impact to a field.

## Discounted Cumulated Fame of Publications

Next we analyze the citation impact of the top-cited publication of each of 11 authors (see Appendix and the section “Publications as Units of Analysis”). Figures 7 – 9 present the unweighted discounted impact of the eleven top publications for three decay values  $b = 1.1$ , 2, and 10. Generally, the curves are convex, and the more so, the faster the decay employed. The (non-dynamic) graphs report the past impact discounted to the year of evaluation (2009).

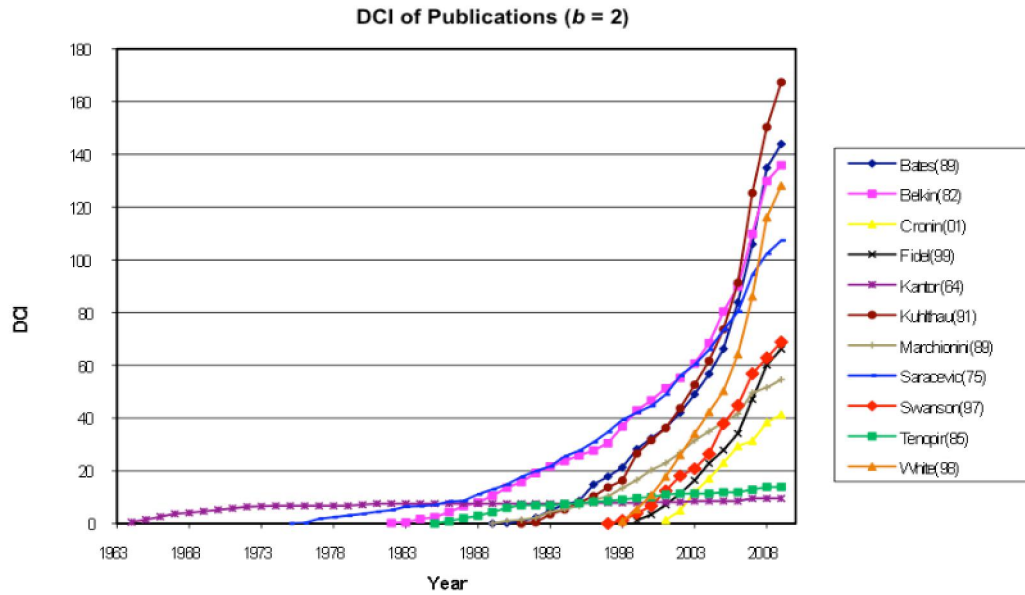


**Figure 7.** DCI of publications with  $b = 1.1$ , no citation weights.

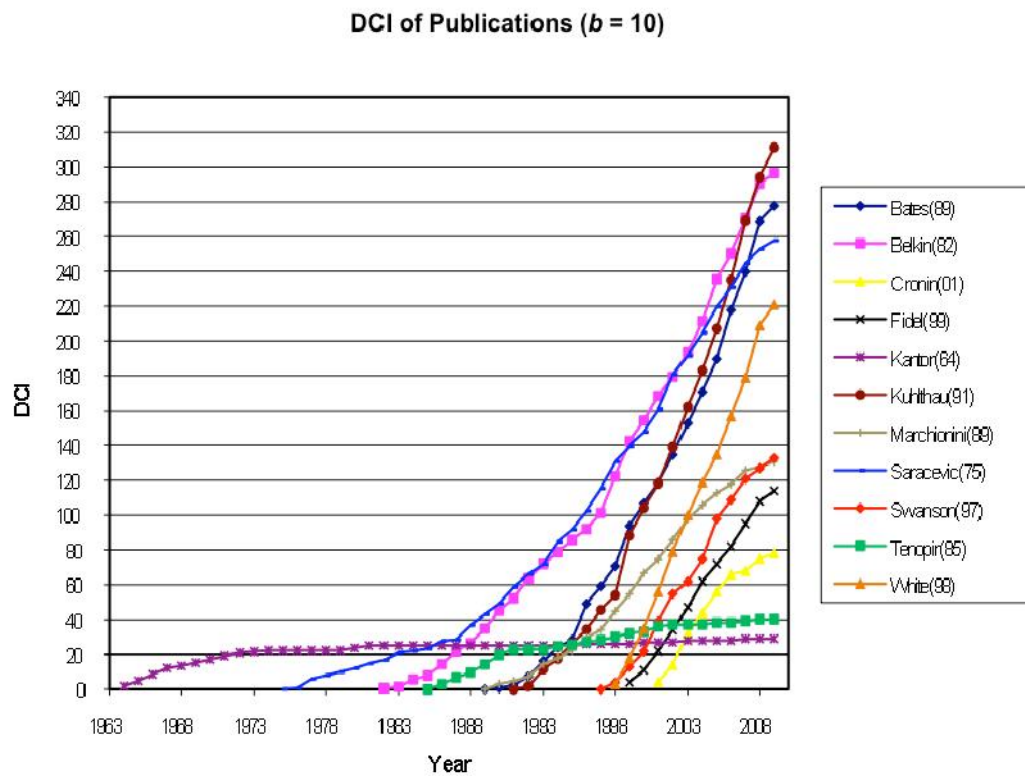
Figure 7, with fast decay ( $b = 1.1$ ), clearly suggests that no publication’s past impact beyond 5 year past in the history is of notable importance today. The end-points of the curves indicate the ‘hottest’ current publications among the eleven. With rapid decay, new publications have an equal opportunity for (historyless) current impact. The publication years range from 1964 to 1999. The currently ‘hottest’ publication is Kuhlthau (1991), followed by White (1998), Bates (1989), and Belkin (1982). Even relatively old papers may retain their status of being continuously cited. Such papers are likely to be seminal papers for entire research traditions, or papers pointing out great challenges, within a discipline.

Figure 8, with a more moderate decay ( $b = 2$ ), changes the currency of the top papers by incorporating a bit more citation history. The curves are less convex and the long-term (over 10 years) impact of Belkin’s and Saracevic’s begins to stand out. While Kuhlthau and Bates keep their strong positions, Belkin and Saracevic (1975) climb up and White (1998) drops slightly.

By hauling over more citation history through a slow decay ( $b = 10$ ), Figure 9, four top publications stand out: Kuhlthau (1991), Belkin (1982), Bates (1989), and Saracevic (1975). For those working within information science, this hardly is surprising. Common knowledge within the field points these publications as belonging to its Hall-of-Fame.



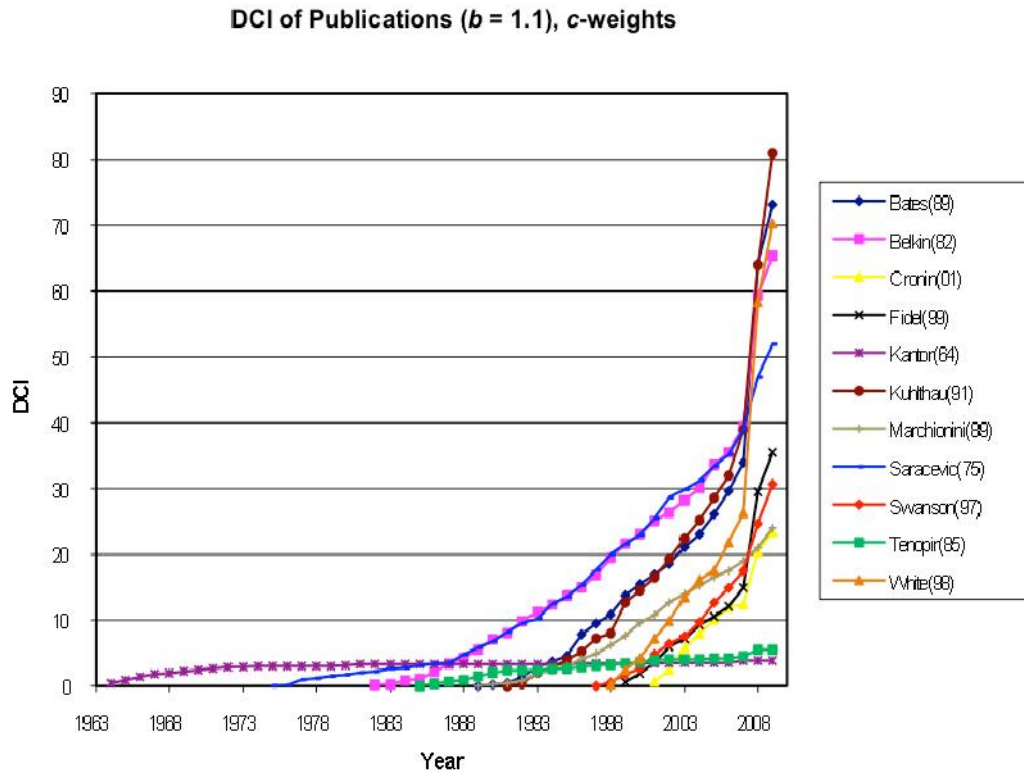
**Figure 8.** DCI of publications with  $b = 2$ , no citation weights.



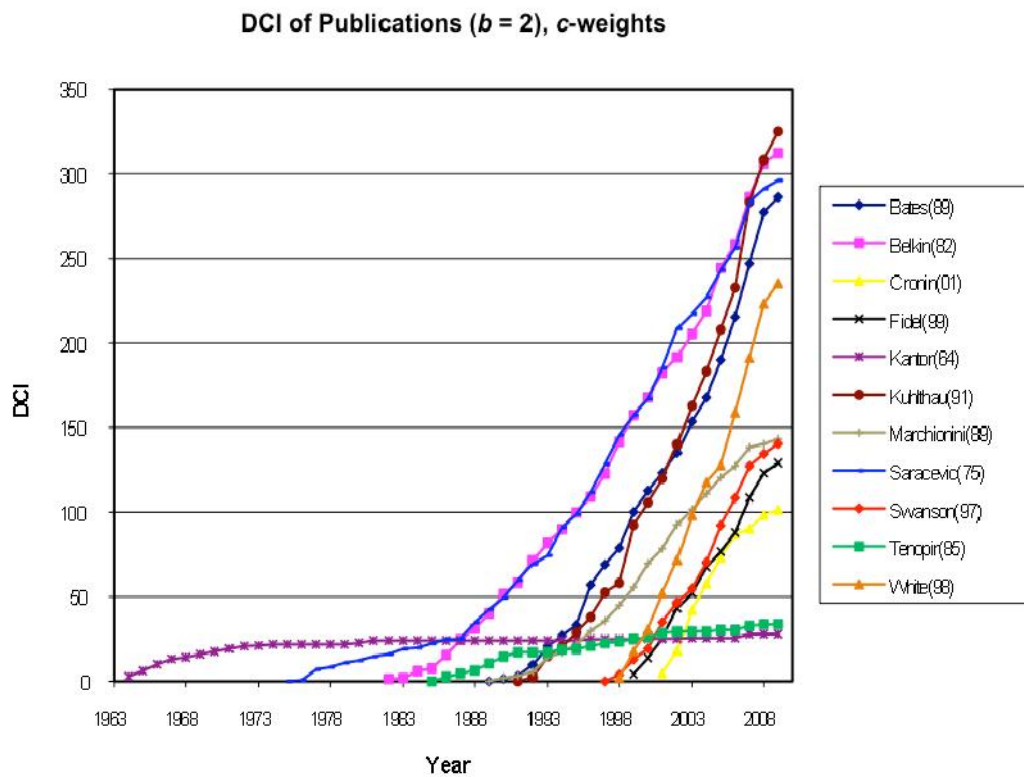
**Figure 9.** DCI of publications with  $b = 10$ , no citation weights.

Figures 10 – 12 present the weighted discounted impact of the eleven top publications for the three decay values:  $b = 1.1$ , 2, and 10. Overall, the curves seem similar to their unweighted counterparts, but there still are some important differences.





**Figure 10.** DCI of publications with  $b = 1.1$  and citation weights.

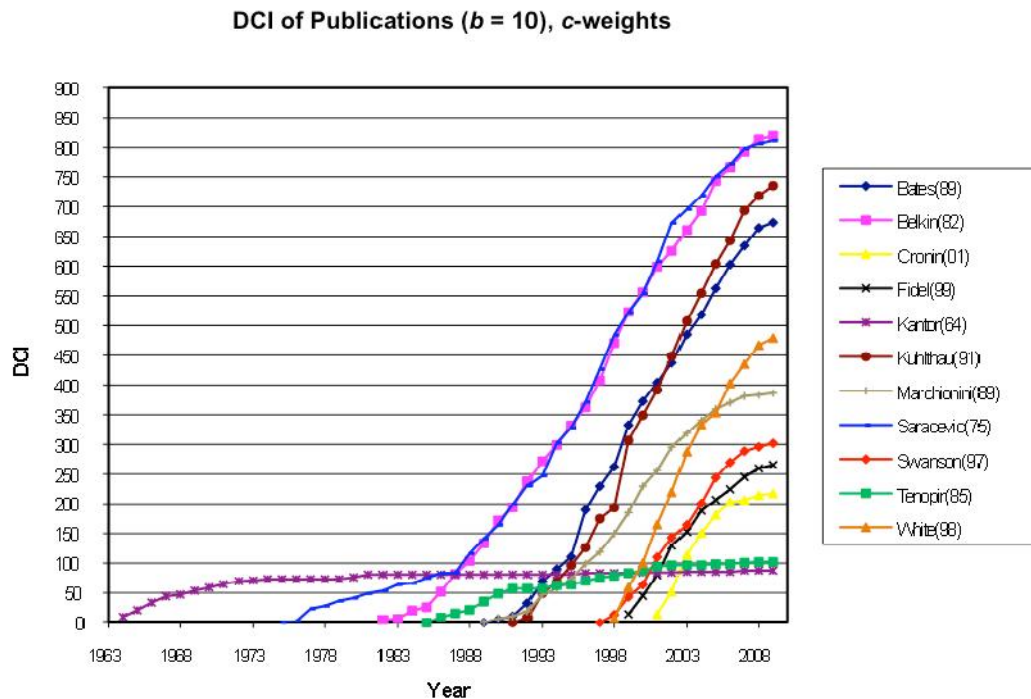


**Figure 11.** DCI of publications with  $b = 2$  and citation weights.

With fast decay ( $b = 1.1$ , Figure 10), the curves are clearly less convex than the unweighted ones. Weighting boosts the discounted historical impact of the top publications (now

excluding White's), suggesting that other high-impact publications have cited them already a long time ago.

By relaxing the decay (Figures 11 and 12), citation weighting clearly boosts the top four publications (by Bates, Belkin, Kuhlthau, and Saracevic – three of them from the same institution!) and also changes their terminal order. Also Marchionini's (1989) publication strengthens its relative position, indicating relatively many citations by other high-impact publications and thus greater impact.

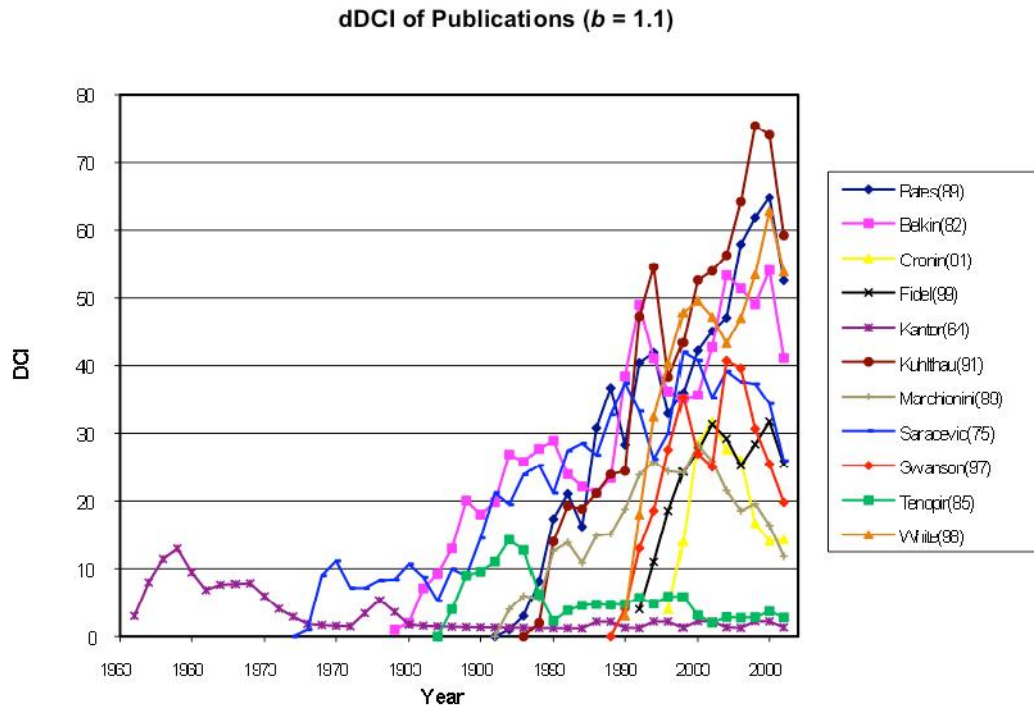


**Figure 12.** DCI of publications with  $b = 10$  and citation weights.

## Dynamic Discounted Cumulated Fame of Publications

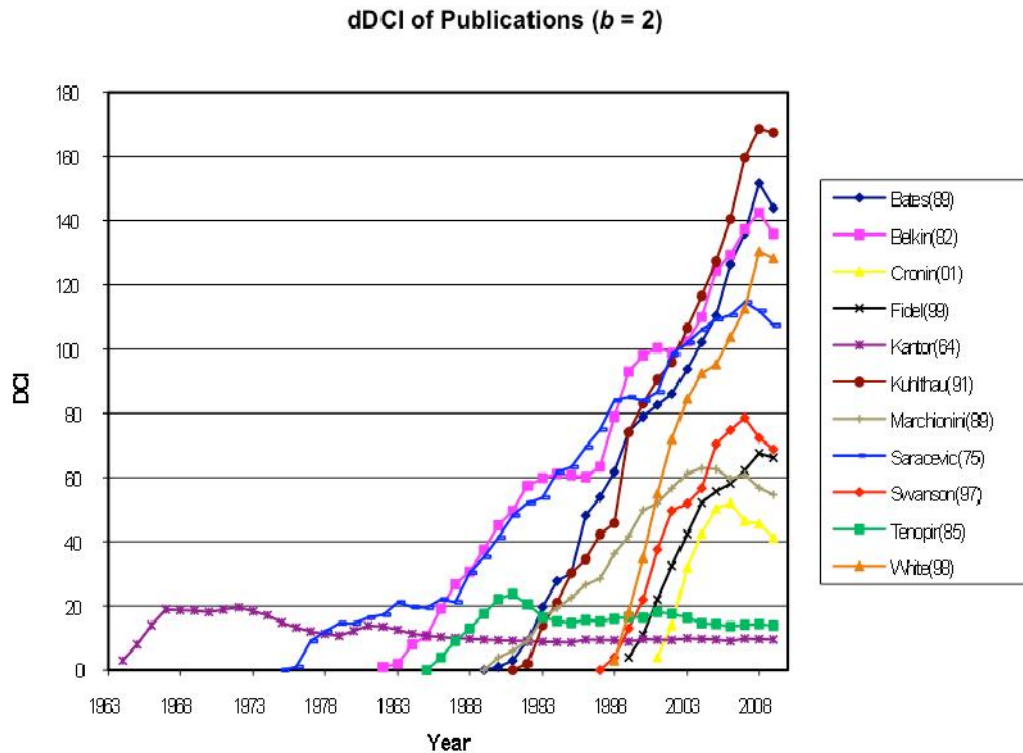
Above we looked at the top publication impact from 1963 to 2009 with the year 2009 as the time point of analysis. Past performance at each year was discounted to the year 2009. We now move to dynamic impact, where the year of analysis and discounting moves across the years as well. Figures 13 – 15 report the unweighted dynamic impact of the 11 top publications with varying decay speed ( $b = 1.1, 2, 10$ ).





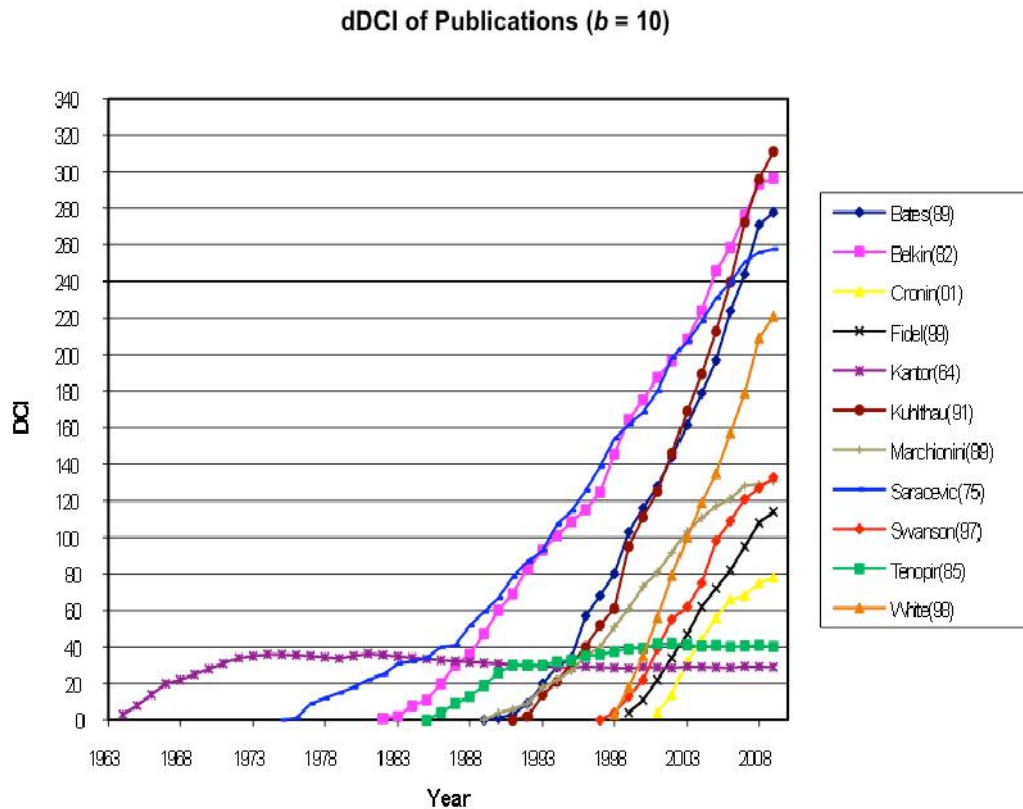
**Figure 13.** Dynamic DCI of publications with  $b = 1.1$ , no citation weights.

With fast decay ( $b = 1.1$ , Figure 13) the curves are again rather jagged reflecting the short-term fluctuations in each publication's impact. Over the years, there are more and more curves as new publications appear. From the beginning of 1990's, the top ranks have much variation. For the last 10 years, Kuhlthau (1991) is the 'hottest' individual publication in almost every year. Bates (1989), Belkin (1982) and White (1998) are also within the top ranks.



**Figure 14.** Dynamic DCI of publications with  $b = 2$ , no citation weights.

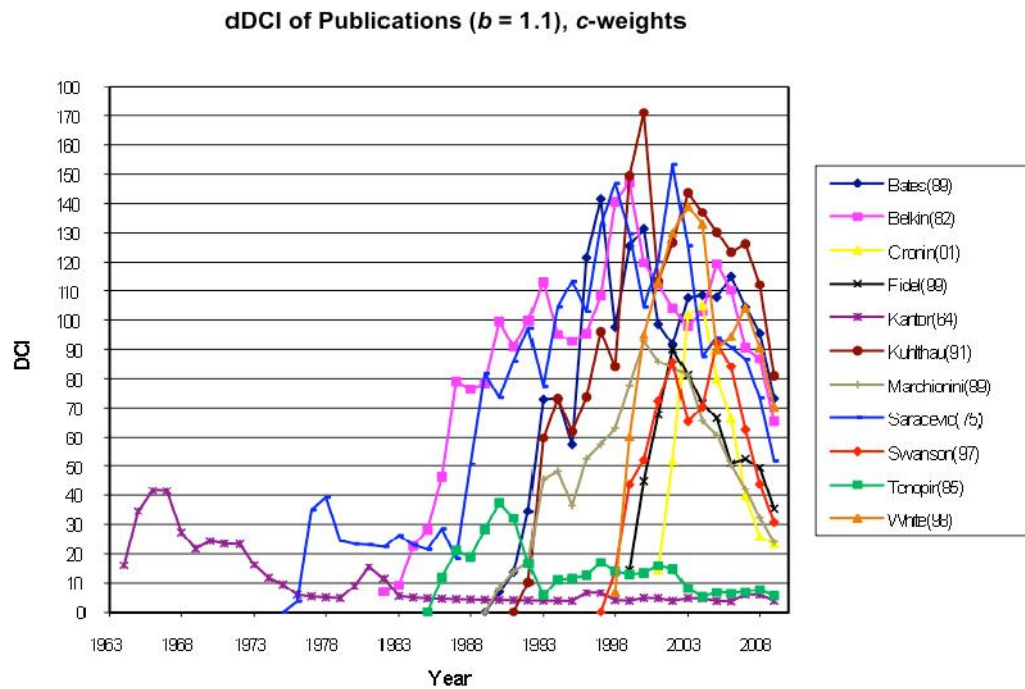
With a more moderate decay ( $b = 2$ ; Figure 14) of citations, one may see Belkin (1982) and Saracevic (1975) strengthening their impact while Kuhlthau (1991) continues at the top for the last years. With slow decay ( $b = 10$ , Figure 15), the four previously identified top publications stand out and White (1998), one of the most recent ones, still climbs close.



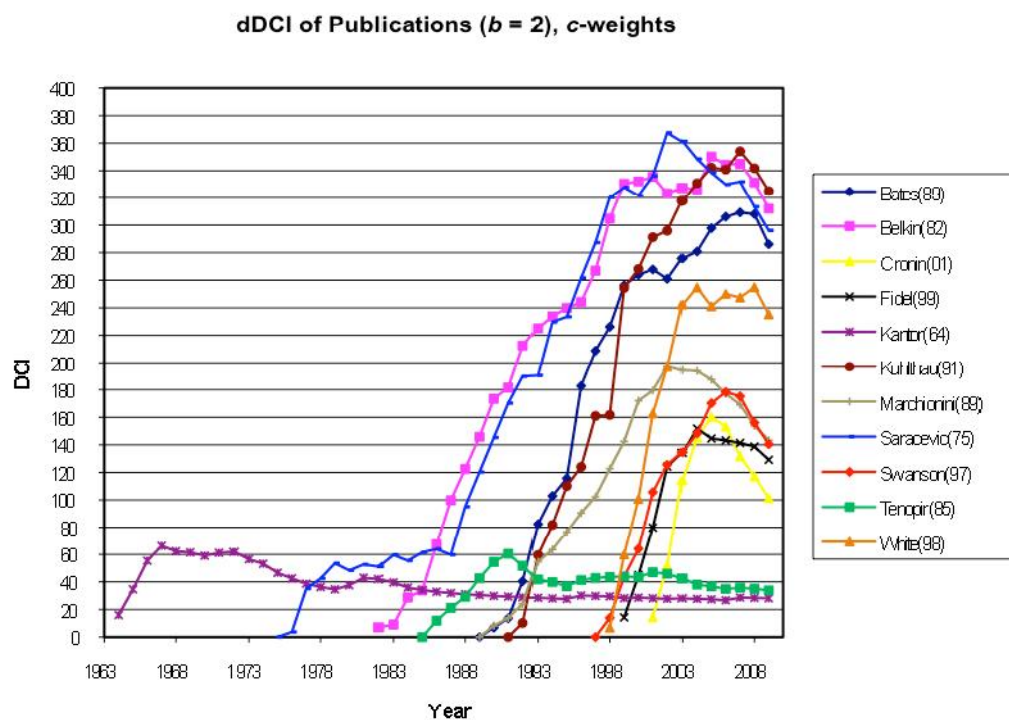
**Figure 15.** Dynamic DCI of publications with  $b = 10$ , no citation weights.

Note that the unweighted dynamic graph with slow decay ( $b = 10$ ; Figure 15) is almost equal to its non-dynamic counterpart (Figure 9). Here the early impact of old publications obtains more credit.

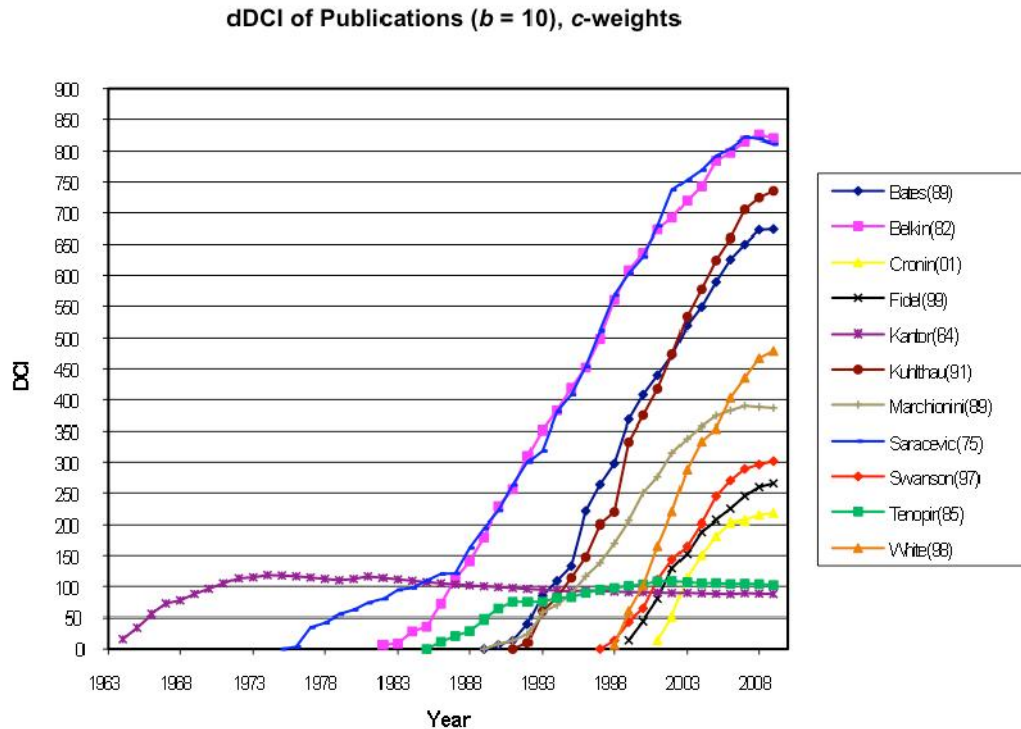
When citation weighting is applied, the overall shape of the curves changes from the unweighted curves: the curves dive for the last years of analysis (except for slow decay). This is a product of the method, because for the last years the citing publications have not had time to gain weight. Nevertheless, this situation is equal across the publications studied. With moderate or slow decay, weighting improves the relative standing of Belkin (1982) and Saracevic (1975) (Figures 14 and 15 compared to Figures 17 and 18, respectively). Across Figures 16 – 18, one may notice that the impact of the top publications, of Belkin (1982) and Saracevic (1975) in particular, is boosted when decay slows down (i.e., a longer impact history is observed in each year). Further, when decay slows down, the overall scores rise, curves become smoother (as above for the unweighted case), and the last years' dive becomes less prominent (because longer-term weighted impact is taken into account).



**Figure 16.** Dynamic DCI of publications with  $b = 1.1$  and citation weights.



**Figure 17.** Dynamic DCI of publications with  $b = 2$  and citation weights.



**Figure 18.** Dynamic DCI of publications with  $b = 10$  and citation weights.

Note that the weighted dynamic graph with slow decay ( $b = 10$ ; Figure 18) is almost equal to its non-dynamic counterpart (Figure 12). Here the early impact of old publications obtains more credit.

## Discussion

Cronin and Meho (2007) analyzed the citation impact of 12 authors who have won either the American Society for Information Science and Technology Award of Merit or the Research in Information Science award. For each author, Cronin and Meho analyzed the authors' overall cumulative citation impact, and the impact of their high-impact publications (year and citations cumulated up to 2006). They also report the author's age and year of PhD along the timeline. Their focus is the relationship between creativity and each author's chronological and professional age.

Our evaluation exercise also reports on the same set of authors and their single most heavily cited publications while we do not look at the relationship of creativity and time. In the case of high-impact publications, we chose just the top-publication for analytical clarity but could of course have analyzed all high-impact publications by the Cronin-Meho standard, i.e., publications which have received at least 40 citations.

In our view, both our analyses, on authors and on publications, shed light on and extend the Cronin and Meho results in important ways. Regarding authors:

- *Decay*: The Cronin and Meho cumulative citation curves rise monotonically; by superimposing them, one may observe that some curves start earlier than others, some rise higher, some rise faster than others at some point. Nevertheless, these graphs treat all citations equal and thus allow a scholar to 'rest on his/her laurels'. Our DCI method treats each equally old citation equal but old citations less rewarding than newer ones. However, the age of the

cited publication does not matter. By controlling the decay parameter, research evaluation can be tuned more history-oriented or more currency oriented. For example, in Figures 1 – 3, one may observe that Swanson's relative position greatly improves when the analysis moves from currency to long-term impact. On the contrary, Marchionini's and Kuhlthau's progress toward top ranks shows as high currency during the most recent years (Figure 1), while their impact still remains less prominent in longer perspective (Figure 3). Belkin's and Saracevic's high standing is apparent irrespective of the perspective: they seem to have continuous currency.

- *Citation weights*: Research evaluation by traditional citation impact treats all citations equal. One may nevertheless reasonably argue that citations by highly cited publications are more important since the latter, transitively, more effectively push forward the impact of the original publications since these citing publications have been frequently cited. Our findings provide indirect support to this view: by comparing the corresponding citation-weighted and unweighted graphs, one may observe the difference caused by weighting: some authors gain impact while some others lose.

- *Normalization*: Research evaluation by traditional citation impact allows author comparison by absolute numbers. For example, one may superimpose the cumulative citation counts by Cronin and Meho (2007) and observe the absolute citation count development over time, the beginning and ascent of individual author's curves etc. However, if there is a change in citation volume in one field over time, or if two fields of quite different citation volume are compared, traditional absolute impact figures are not fair. The normalized figures (for authors) proposed here give credit to top researchers of a research area with a small citation volume who would have no chance of receiving the citation counts a middle-rank researcher in an area with a large citation volume easily obtains. In our data set, the research and citation volumes grow over the years. Even if some scholar is receiving increasing numbers of citations, also the baseline (average for the field) may grow as well. Our normalized dDCI curves make relative impacts comparable.

Regarding publications:

- *Decay*: Cronin and Meho (2007) analyze, which publications by the award-winning authors have had high impact and at which stage in the author's career they were published. We think that this is of great interest in the analysis of the relationship of impact and stage of career. However, we also think that it is important to be able to analyze the life span of such high-impact publications – whether they were short-lived or have sustained interest. Our (d)DCI curves of publications assess the current (DCI) and dynamic historical (dDCI) impact of key publications. For example in 2009, Kuhlthau (1991), White (1998), Bates (1989), and Belkin (1982), in this order, are the key publications regarding currency ( $b = 1.1$ , see Figure 7). Regarding longer-term impact, the order changes and other publications (e.g., Saracevic 1975) challenge the ones mentioned (Figure 9). The dynamic impact curves of the publications (Figures 13 – 18) show their rise and fall from the perspective of each year of evaluation. For example, Figure 13 and 14 indicate that Kantor (1964) and Tenopir (1985) peaked in impact in late 1960's and late 1980's (respectively) where after their impact has diminished.

- *Citation weights*: Like with authors, citation weighting in the analysis of individual publications helps to identify publications contributing to later high-impact publications. A comparison of Figures 7 – 9 vs. 10 – 12 shows that citation weighting promotes some

publications (e.g., Bates 1989, Belkin 1982, and Saracevic 1975 between Figures 7 and 10) at the cost of other publications.

- *Normalization*: Our findings do not cover normalized impact of individual publications, since the normalized comparison of top publications across authors seems less interesting. However, normalized dDCI analyses across the publications of an individual author would be of interest and could be performed with the proposed methods – but are excluded here.

These two levels of aggregation in research evaluation, individual publications and individual authors, suggest by induction, that other higher aggregation level objects such as research institutions, universities, disciplines, or countries, could be subject to the proposed DCI-based evaluation methods.

If evaluation concerns, say, an author, only in relation to the current year, not only the DCI but also the dDCI reveals the requested citation information. However, DCI is simpler and sufficient if past standing of the objects under evaluation is of no interest. It brings the impact to the point of evaluation but does not show the rise and fall (if any) of the author's impact while dDCI conveys this.

In addition to the strengths of the DCI-based evaluation methods discussed above, there are some issues that one needs to consider when applying them.

Firstly, whether the findings on rising and falling stars correlate with other measures of scholarly performance, e.g., peer assessments, remains to be seen. Even if the results would suggest interesting hypotheses, the numbers themselves do not carry explanations, which may vary from individual performances to general development of each author's subdomain in a domain. The DCI-index, like other bibliometric indicators, should complement, and not replace, sound judgement.

Secondly, regarding the specific data set analyzed in the present paper, one should bear in mind that while the normalized author impact curves show rising and falling stars, each of the 12 authors belong to the cream-of-the-cream of information science, above the many others that publish in our top journals and conferences.

Thirdly, regarding normalization, it matters which set objects are analyzed. This is not a trivial choice and must be taken into account in the interpretation of results. The results would be different, if the author set would contain other researchers in information science, e.g., those working mainly in information retrieval.

Fourthly, like in all informetric studies, the data set (its limitations) greatly affects the findings. If our data set for the 12 authors came from Google Scholar instead of the Web of Science, the findings would probably be different.

We believe, that by bearing the above issues in mind, the DCI-index is a valuable tool when one wants to weight the relative positions of established scientists in a field, identify new promising researchers for research funding, track the impact of key papers in a field, or identify the best papers of sustained value or current influence published in a journal or conference.

## Conclusions

In the present paper, we reanalyzed an earlier data set by Cronin & Meho (2007) on American Society for Information Science and Technology Award of Merit or the Research in Information Science Award winners, with the aim of exemplifying the salient features of the DCI-index proposed earlier by Järvelin & Persson (2008). We applied the index on, and compared the results to the outcomes of, the Cronin & Meho (2007) study. We used both authors and their top publications as units of analysis, and showed that, by adjusting the parameters of evaluation according to current needs of research evaluation, the DCI index delivers data on:

- a units's 'lifetime' impact or current impact at the time of evaluation
- a units's capability of attracting citations from highly-cited later publications as an indication of impact, and
- the normalized relative impact of across a set of units over their 'lifetime' or currently.

Let  $A$  be a body performing research evaluation and intending to use the DCI index. If  $A$  is interested in overall impact of the unit of analysis,  $A$  should use large value of the decay parameter  $b$ . If, on the contrary,  $A$  is only interested in current impact, a small value on  $b$  should be employed. If  $A$  cares about the impact of citing publications, then the 2<sup>nd</sup> generation citations should be taken into account. If  $A$  is interested in the relative standing (with respect to the data set) of the unit of analysis, the normalized (d)DCI should be used.

The DCI-index can reveal rapidly rising new star units in a domain as well as units that are becoming part of the history of a domain in the sense of no more being actively cited. Traditional analyses based on total accumulated citations, or the  $h$ -index, do not directly support this, and require special treatment of the data sets used.

The publication analysis clearly pointed out top-class publications among several excellent ones. While the identification of those publications might not surprise knowledgeable scholars in information science, in research evaluation the evaluators do not, as a rule, have such domain-specific knowledge. Thus, the DCI-index provides a useful tool for verifying any claims on impact in a field.

## Acknowledgements

The authors thank Lokman Meho for making the citation data set on ASIST Award winners used in Cronin & Meho (2007) available. This research was funded, in part, through the Academy of Finland Grants #120996 and #124131.



## References

- Braun, T., & Glänzel, W. (1990). United Germany: The new scientific superpower? *Scientometrics*, 19(5-6), 513-521.
- Braun, T., Glänzel, W., & Schubert, A. (2005). A Hirsch-type index for journals. *Scientist*, 19(22), 8.
- Burrell, Q.L. (2007). Hirsch index or Hirsch rate? Some thoughts arising from Liang's data. *Scientometrics*, 73(1), 19-28.
- Cronin, B., & Meho, L. (2006). Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, 57(9), 1275-1278.
- Cronin, B., & Meho, L. (2007). Timelines of creativity: A study of intellectual innovators in information science. *Journal of the American Society for Information Science and Technology*, 58(13), 1948-1959.
- Egghe, L. (2006). How to improve the h-index. *Scientist*, 20(3), 15.
- Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch-index. *Scientometrics*, 69(1), 121-129.
- Glänzel, W., Debackere, K., & Meyer, M. (2008). 'Triad' or 'tetrad'? On global changes in a dynamic world. *Scientometrics*, 74(1), 71-88.
- Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569-16572.
- Jin, B.H., Liang, L.M., Rousseau, R., & Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, 52(6), 855-863.
- Järvelin, K., & Persson, O. (2008). The DCI index: Discounted cumulated impact-based research evaluation. *Journal of the American Society for Information Science and Technology*, 59(9), 1433-1440.
- Meho, L.I., & Rogers, Y. (2008). Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 59(11), 1711-1726.
- Moed, H.F. (2005). *Citation Analysis in Research Evaluation*. Dordrecht: Springer.
- Oppenheim, C. (2007). Using the h-index to rank influential British researchers in information science and librarianship. *Journal of the American Society for Information Science and Technology*, 58(2), 297-301.
- Sandström, U., & Hällsten, M. (2008). Persistent nepotism in peer-review. *Scientometrics*, 74(2), 175-189.
- Schubert, A., & Glänzel, W. (2007). A systematic analysis of Hirsch-type indices for journals. *Journal of Informetrics*, 1(3), 179-184.
- Takeda, Y., Mae, S., Kajikawa, Y., & Matsushima, K. (2009). Nanobiotechnology as an emerging research domain from nanotechnology: A bibliometric approach. *Scientometrics*, 80(1), 23-38.
- Visser, M.S., & Nederhof, A.J. (2007). Bibliometric study of the Uppsala University, Sweden, 2002-2006. In J. Nordgren (Ed.), *Quality and Renewal 2007: An Overall Evaluation of Research at Uppsala University 2006/2007*. Uppsala: Uppsala University.
- Wennerås, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, 387(6631), 341-343.
- Zhou, P., Thijs, B., & Glänzel, W. (2009). Is China also becoming a giant in social sciences? *Scientometrics*, 79(3), 593-621.

## Appendix The 11 selected top publications

Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5), 407-424.

Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982). Ask for information-retrieval: 1. Background and theory. *Journal of Documentation*, 38(2), 61-71.

Cronin, B. (2001). Bibliometrics and beyond: Some thoughts on web-based citation analysis. *Journal of Information Science*, 27(1), 1-7.

Fidel, R. et al. (1999). A visit to the information mall: Web searching behavior of high school students. *Journal of the American Society for Information Science*, 50(1), 24-37.

Kantor, P.B. (1964). Nucleon-nucleon scattering + meson resonances. *Physical Review Letters*, 12(2), 52-54.

Kuhlthau, C.C. (1991). Inside the search process - information seeking from the users perspective. *Journal of the American Society for Information Science*, 42(5), 361-371.

Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science*, 40(1), 54-66.

Saracevic, T. (1975). Relevance - review of and a framework for thinking on notion in information-science. *Journal of the American Society for Information Science*, 26(6), 321-343.

Swanson, D.R., & Smalheiser, N.R. (1997). An interactive system for finding complementary literatures: A stimulus to scientific discovery. *Artificial Intelligence*, 91(2), 183-203.

Tenopir, C. (1985). Full text database retrieval performance. *Online Review*, 9(2), 149-164.

White, H.D., & McCain, K.W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*, 49(4), 327-355.